Ames Research Center

An Agent-Based Interface for Ecological Forecasting

Keith Golden,

Ramakrishna Nemani,

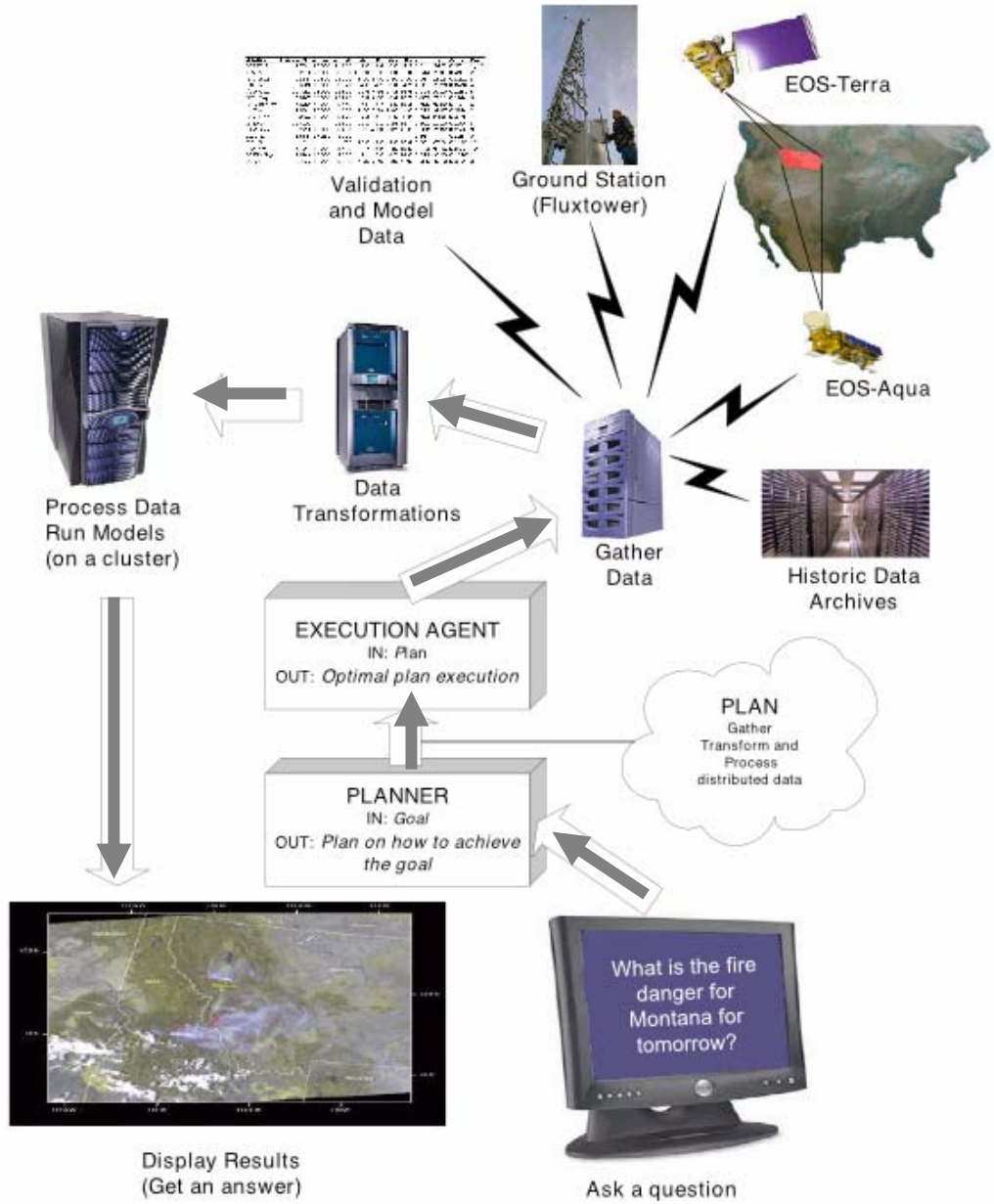
Wanlin Pang,

Petr Votava,

NASA Ames

Oren Etzioni,

U. Washington



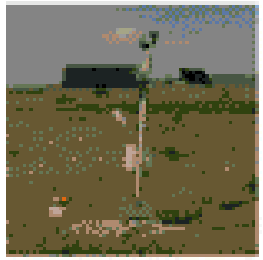
6/22/2004

Agents

- An agent is an intelligent assistant
  - E.g., travel agent
- Provides *goal-oriented* interface
  - You say **what** to do, not **how** to do it.
  - Agent has the knowledge to figure out **how**.
- Copes with uncertainty and error robustly.
  - Obtains information needed to complete task
  - Tries something else when encountering failure
  - Asks for clarification if a request is ambiguous

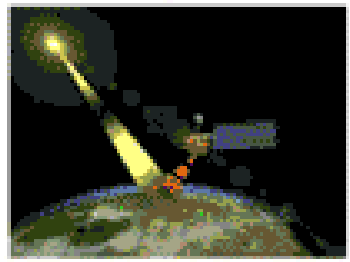
Terrestrial Observation and Prediction System

Weather Networks



Temperature/rainfall/
radiation/humidity/wind

Orbiting Satellites
Terra/Aqua/Landsat/Ikonos

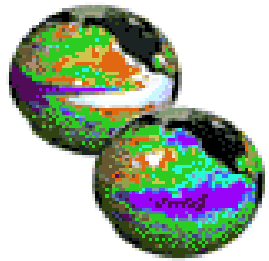


Landcover/
change, Leaf
area index,
surface
temperature,
snow cover and
cloud cover

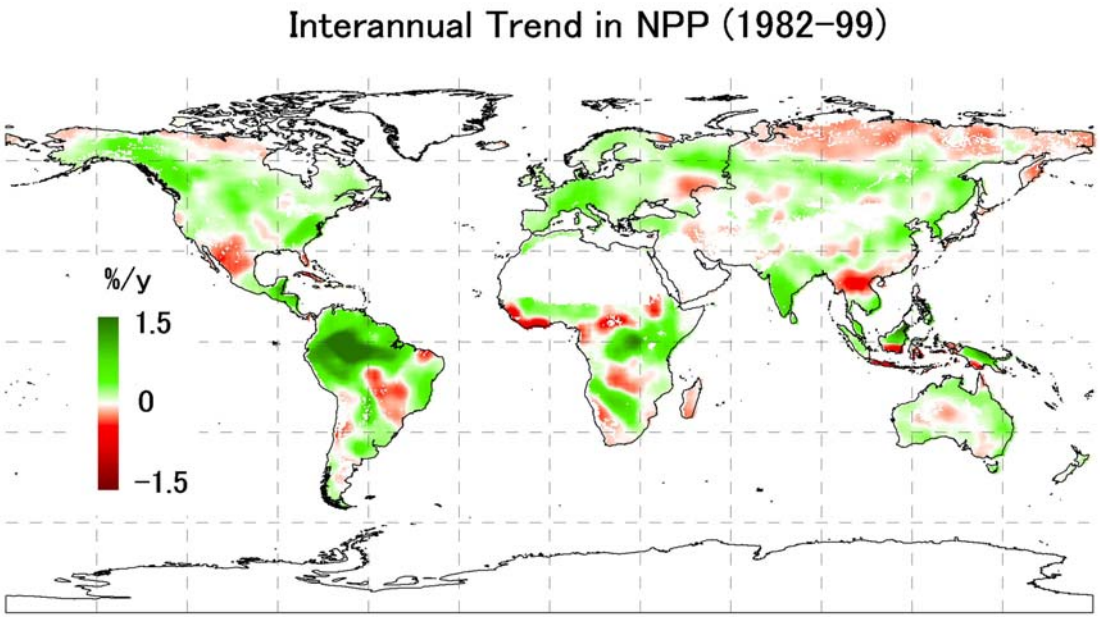
Ancillary Data
Topography, River
networks, Soils

Ecosystem
simulation models

Weather & Climate
Forecasts



Monitoring & Forecasting
Stream flow, soil moisture, phenology, fire risk, forest/range/crop production



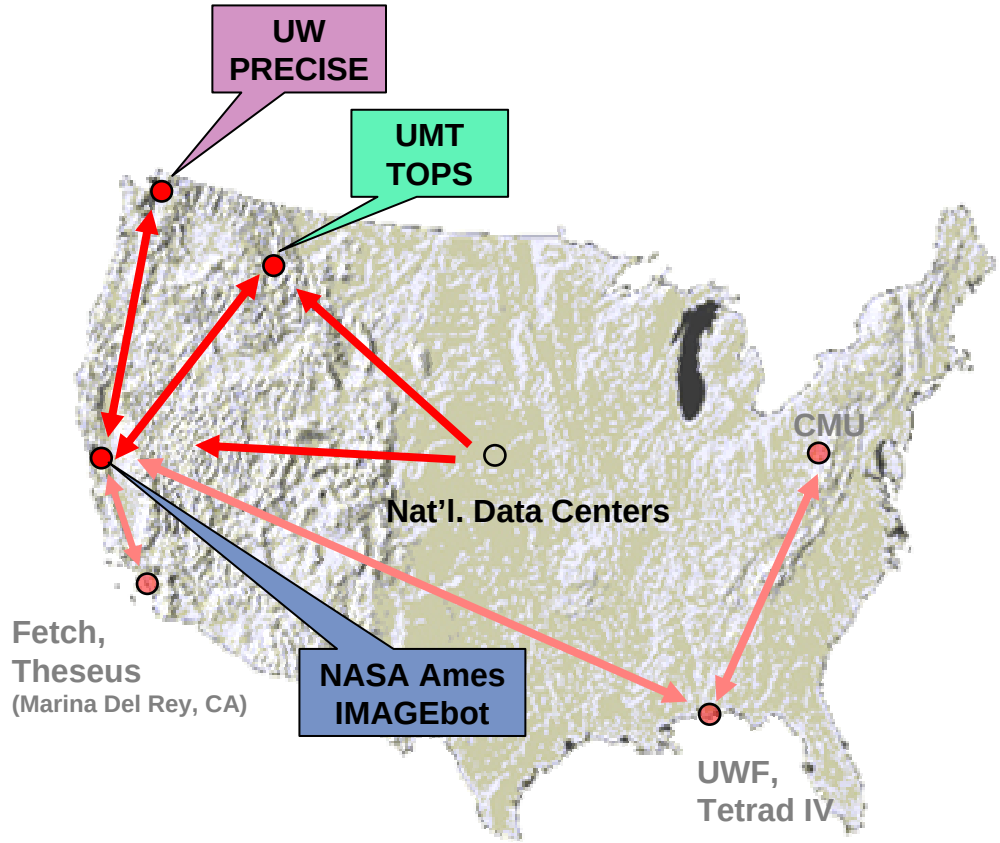
- Climate-driven increases in global terrestrial net primary production from 1982 to 1999. *Science*, 300, 1650 (2003).
  - 3 research assistants for 12 months
  - processed <15 GB of data
  - data preparation >80%

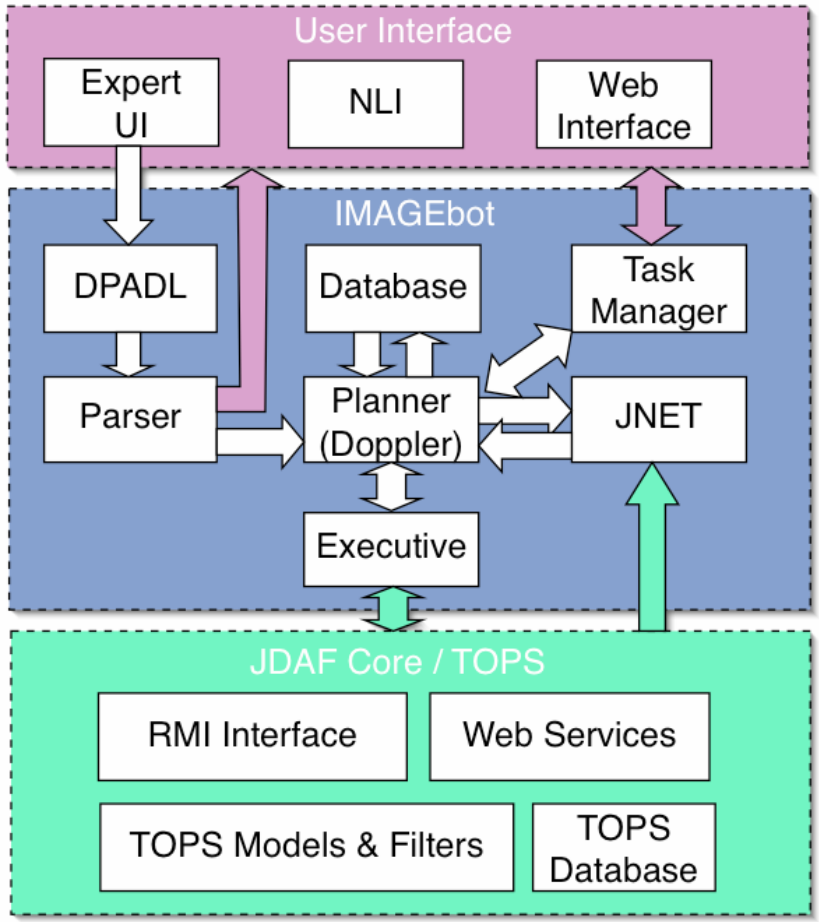
- EOSDIS:
  - Generates ~3 tera-bytes of data a day.
  - Currently holds 2 peta-bytes
  - 1 day = 2 years of HUBBLE Space Telescope

Dimensions of Autonomy

- Automation
  - Automate the generation of the forecasts, analysis of the results, and model adjustments
  - Generate and record “meta-data” to facilitate later searches
- Flexibility
  - Ease of integration of new models and data sources (“plug and play”)
- Robustness
  - Adapt to changes and recover form failures
  - Out of several sources of the same data use the “best” available one

Distributed Agent Architecture



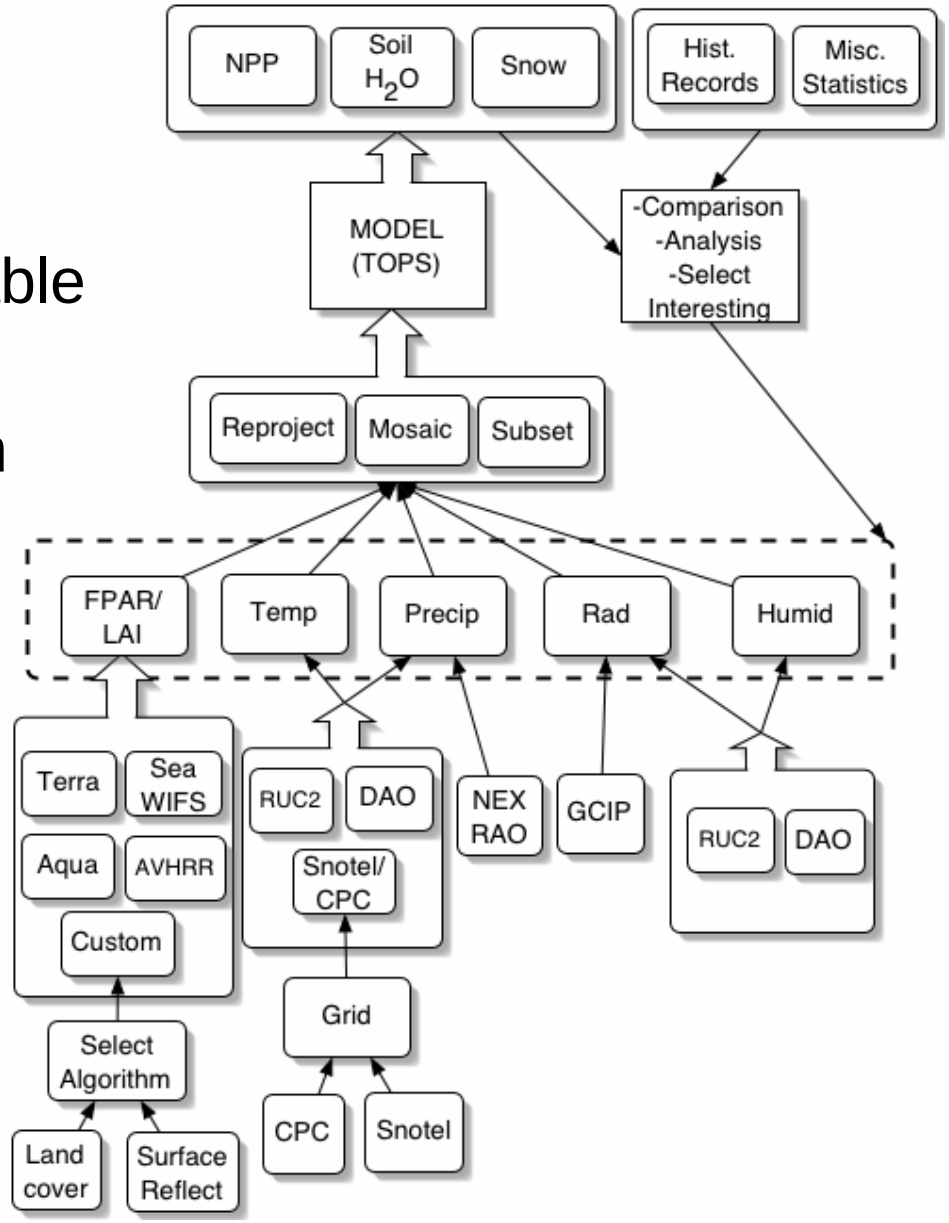


DPADL

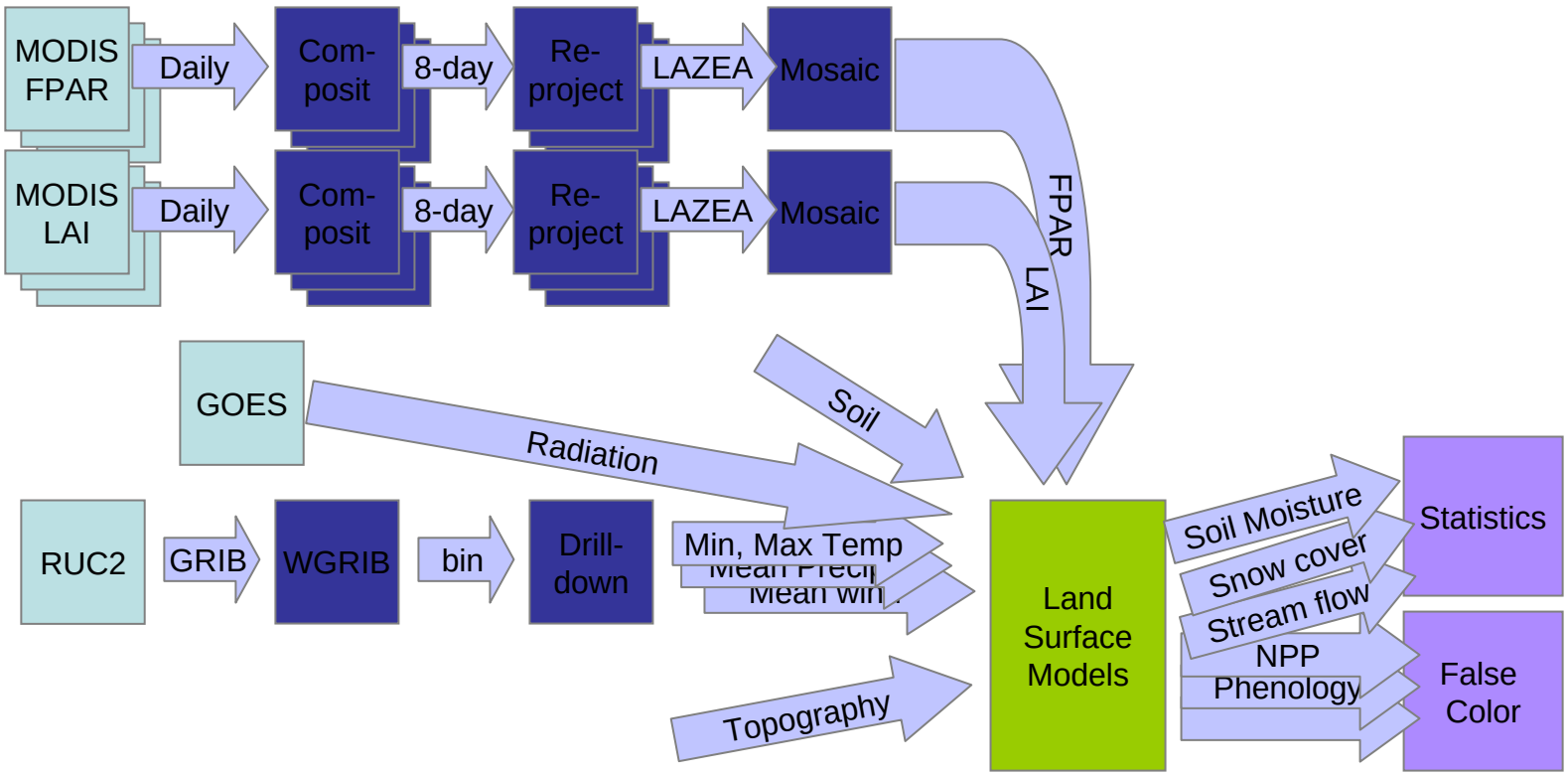
- Inferring effects of data transforms on data requires rich descriptions of both
- Data Processing Action Description Language
- Actions describe data-processing operations
  - Any number of inputs and outputs
  - Causal, declarative representation of data filters
  - Object creation, copying, modification
- Declarative and object oriented
- Arbitrary constraints over any static type
- Integration with Java
  - Embedded Java code
    - Action execution
    - “Procedural” constraints
  - Parameters include Java objects

Planning for data processing

- Initial state = data available
- Goal = data to produce
- Plan = dataflow program
- Actions =
  - Earth system models
  - Data transformations
- Domain characteristics
  - Very large universes
  - Complex data structures
  - Lots of constraints
  - \+ Highly parallel



Dataflow plans

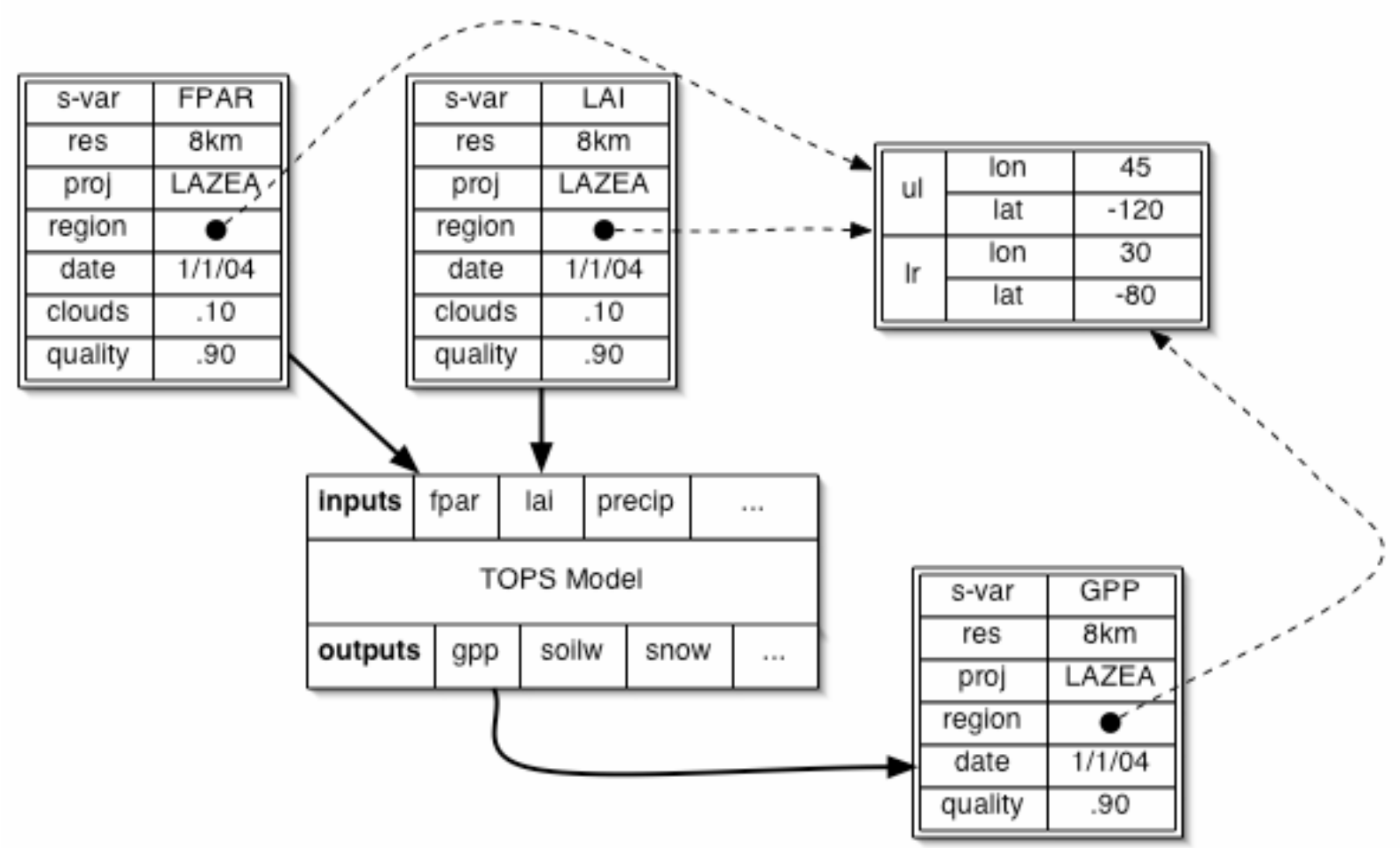


Inputs Filters Models Visualization

Input data choices

Terra-MODIS	FPAR/LAI	1 day	1km, 500m, 250m	global, since 2000
Aqua-MODIS	FPAR/LAI	1 day	1km, 500m, 250m	global, since 2002
AVHRR	FPAR/LAI	10 day	1 km	global, since 1981
SeaWIFS	FPAR/LAI	1 day	1xm x 4 km	global
DAO	temp, precip, rad, humid	1 day	1.25 deg x 1.0 deg	global, since 1980
RUC2	temp, precip, rad, humid	1 hour	40 km	USA
CPC	temp, precipication	1 day	point data	USA
Snotel	temp, precipitation	1 day	point data	USA
GCIP	radiation	1 day	0.5 deg	Continental
NEXRAD	precipitation	1 day	4 km	USA

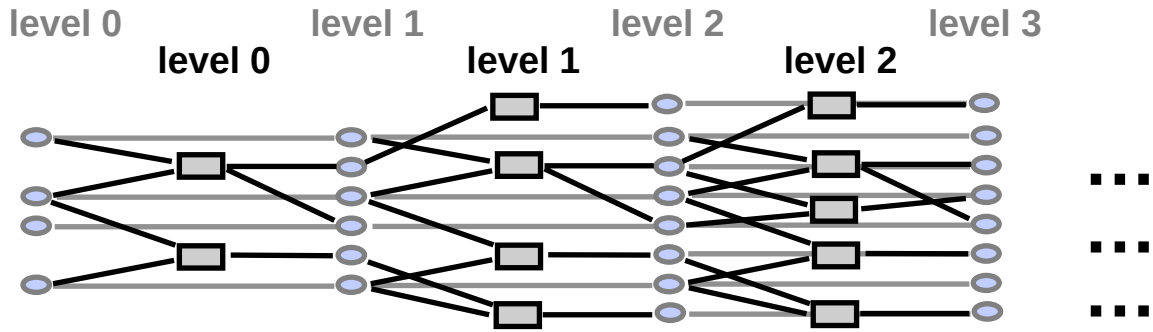
Complex Data Structures



Planning approach

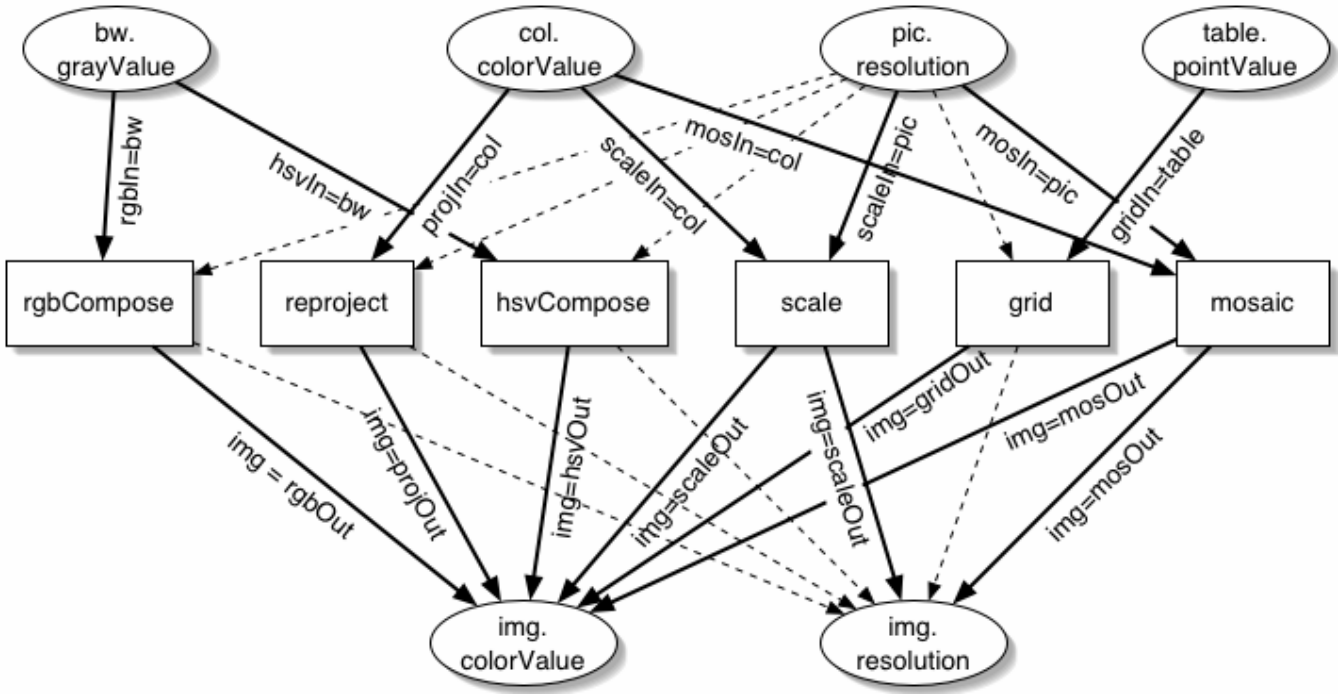
1. Perform graph analysis to derive
   - *distance heuristics*
   - *initial variable domains*
2. Convert planning problem into CSP
3. Search for solution using heuristics
   – Sensors represented as constraints
4. Execute plan, update database and replan if needed

Planning graphs

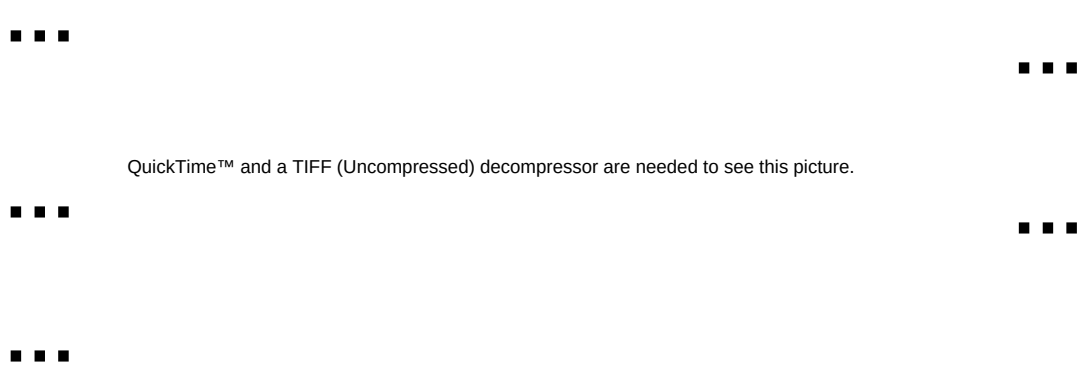


+ Good data structure for reachability/cost analysis
  + Polynomial in problem size
  + Cost = number of steps until proposition(s) possible
– Propositional representation
  – All possible ground actions
  – All possible ground propositions
  – Not feasible when number of objects is large
    - thousands of objects could lead to trillions of mutexes

Lifted planning graphs

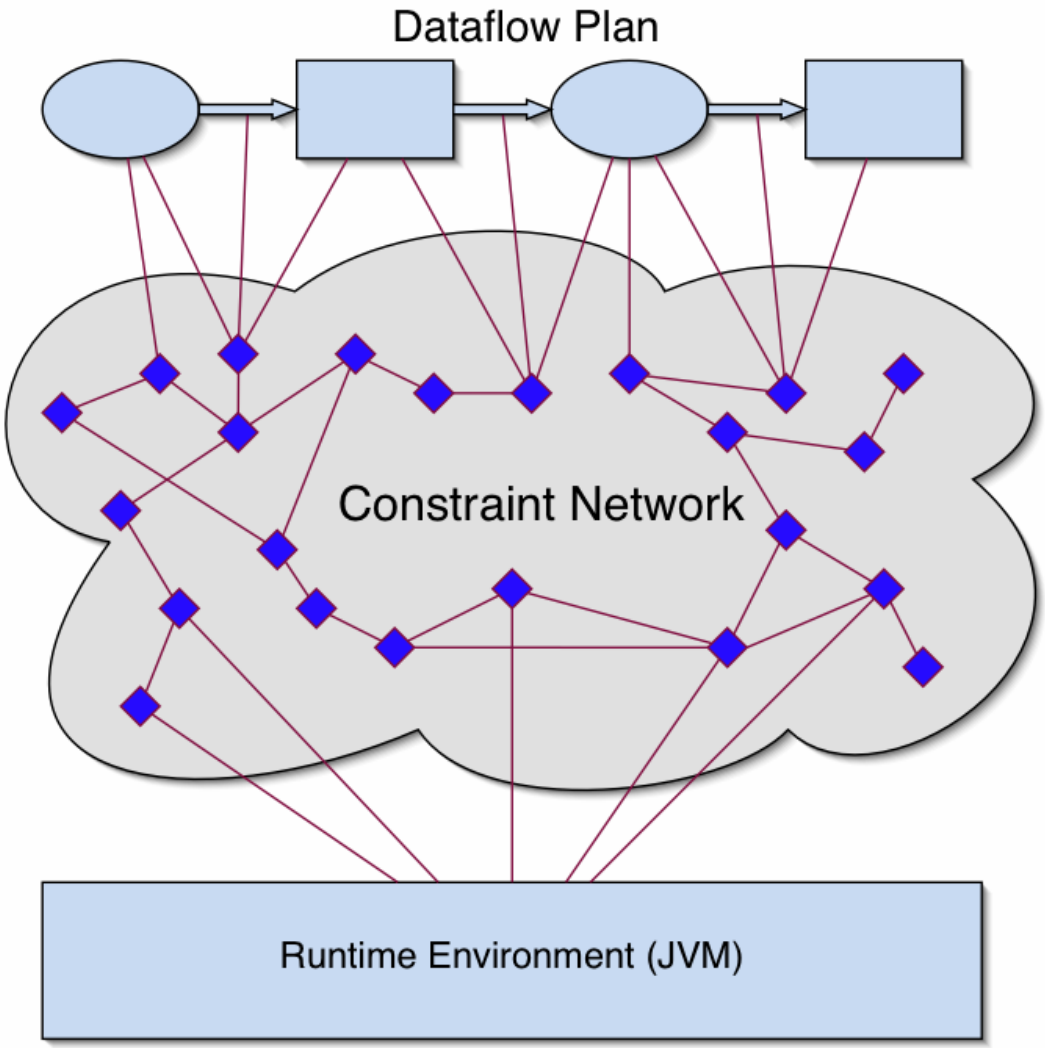


Constraint propagation in PG



- Variable domains recorded for each node
- Values propagated along arcs in graph
- Choice/disjunction → union domains
- Conjunction → intersect domains
- Node splitting to improve propagation

Constraints as glue



JDAF

- Client-Server
- Provides core services for
  - Data acquisition
  - Algorithm execution
- Supports many data types, ESML
- Exploits parallelism, provides load balancing
- Written in Java
- Simple API
- Exposed through both RMI and Web Services

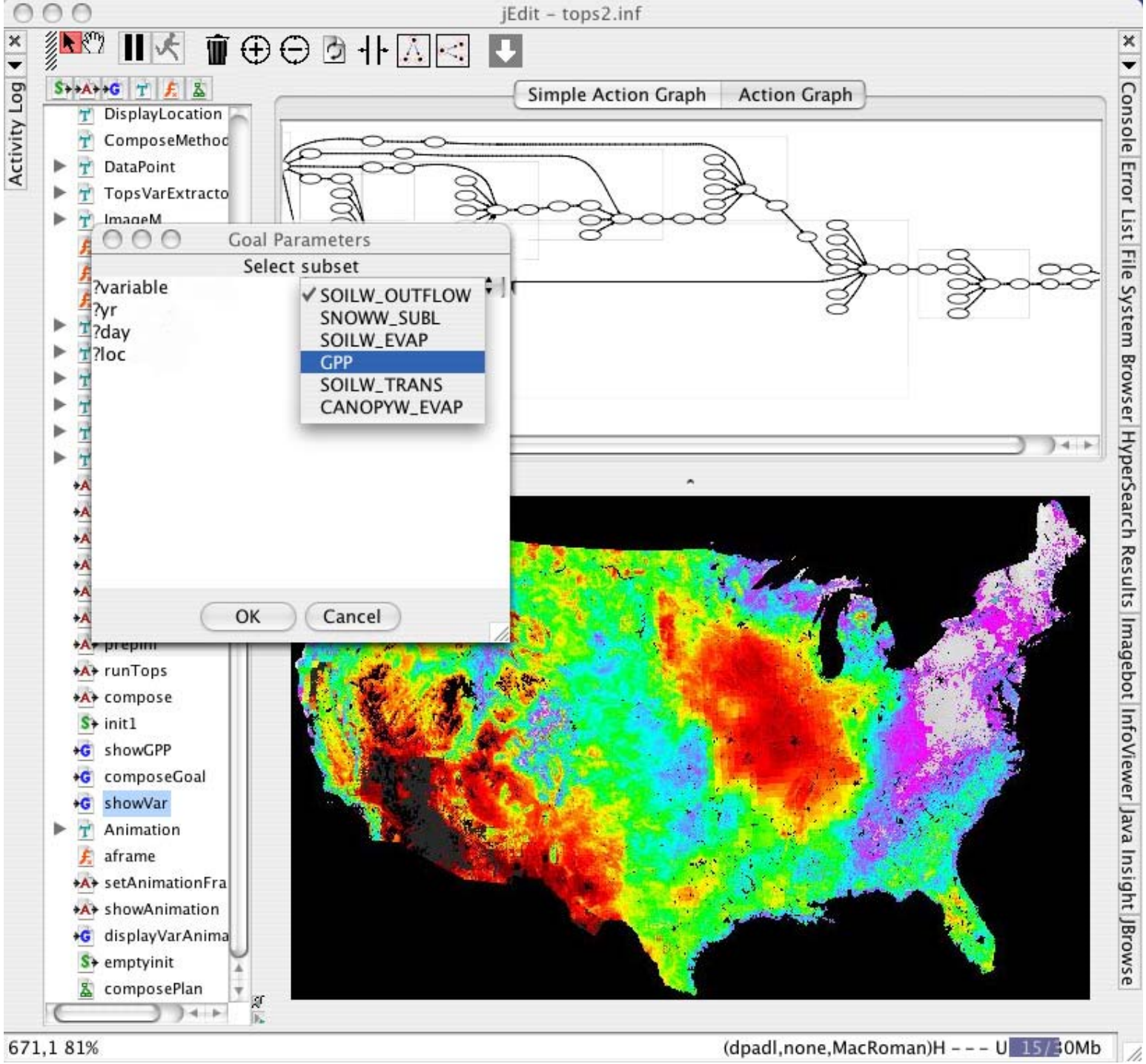
Data Goals and Metadata

- Data product specification
  - **What** information is contained
  - **How** information is encoded in data
  - **Where** the data files are stored/delivered
  - **When** the information pertains to
- Examples
  - I want an MPEG movie of yesterday's weather over the SF bay placed on our website
  - File dd010101.tar.gz is a compressed archive of the downlink directory as of Jan 1, 2001

What
How
Where
When

JDAF Processing Flexibility

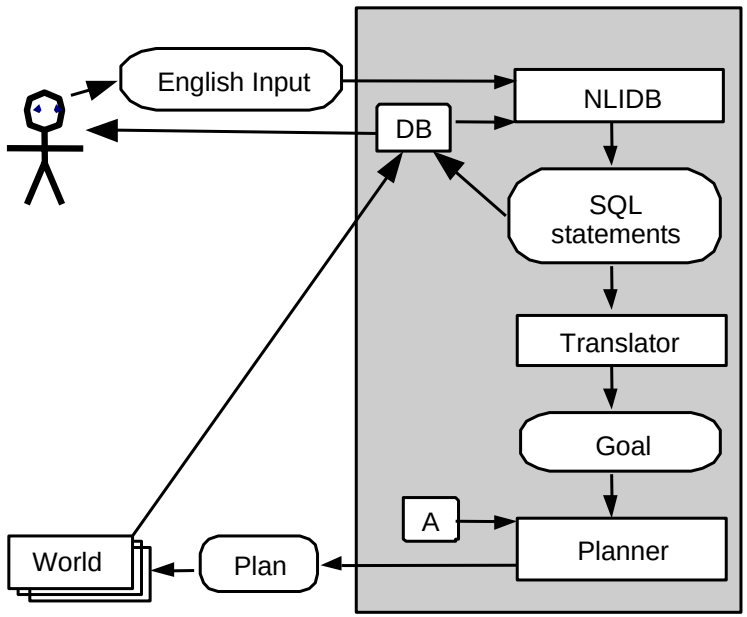
- Adding a new processing algorithm
  - Write/modify and compile
  - Put it to place where the JVM can load it
  - No changes on the server side
- Need only implement simple interface
  - Needed by the server for execution
  - Must implement the execute() method
  - Easy integration even with C/C++ native code
- Schedulable tasks


jEdit - tops2.inf
Activity Log
Simple Action Graph
Action Graph
DisplayLocation
ComposeMethoc
DataPoint
TopsVarExtracto
ImageM
Goal Parameters
Select subset
?variable
?yr
?day
?loc
SOILW_OUTFLOW
SNOWW_SUBL
SOILW_EVAP
GPP
SOILW_TRANS
CANOPYW_EVAP
OK
Cancel
runTops
compose
init1
showGPP
composeGoal
showVar
Animation
aframe
setAnimationFra
showAnimation
displayVarAnima
emptyinit
composePlan
Console
Error List
File System Browser
HyperSearch Results
Imagebot
InfoViewer
Java Insight
JBrowse
671,1 81%
(dpadl,none,MacRoman)H - - - U 15/30Mb

Natural Language Interface

Build on Precise NLIDB

1. English → SQL
2. Resolve ambiguity
3. SQL → DPADL
4. Resolve conflicts
5. Planner solves goal
6. User updated with result



PRECISE queries

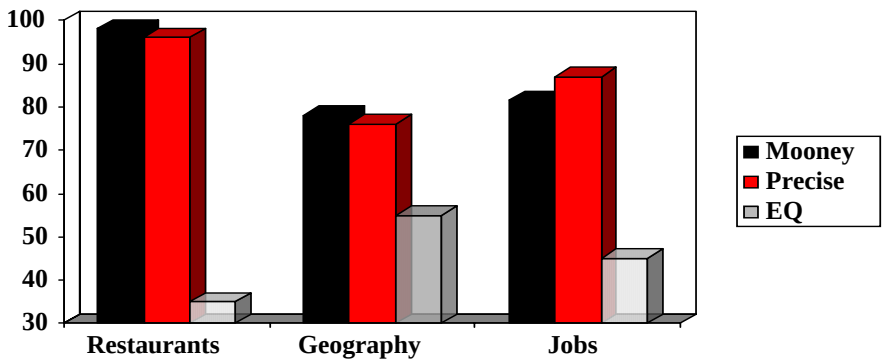
- What is the area of Alaska?
- What is the population of New York?
- Which are all the states that border Oregon?
- How many major cities are in Florida?
- How long is the Colorado River?
- What rivers traverse Indiana and Illinois?
- What rivers traverse Indiana or Illinois?
- What cities are in Texas and have a population of less than 100000 people?
- What is the largest city in the smallest state in the US?
- What states border the state with the largest population?

http://www.cs.washington.edu/research/projects/WebWare1/www/precise/precise.html

Comparison to state of art

Fraction Answered

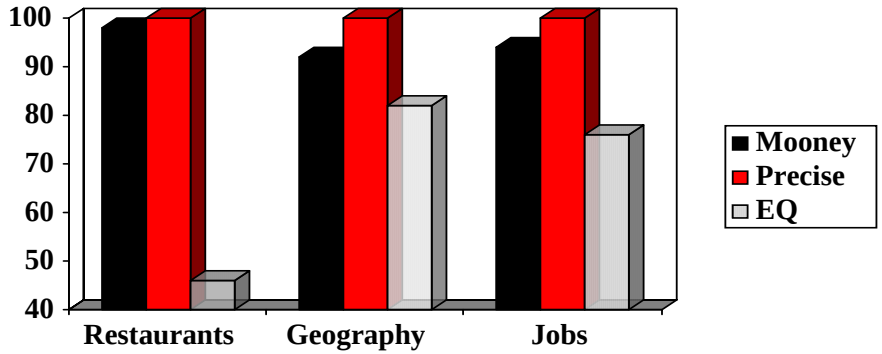
Recall = $Q_{answered}$ / Q



Recall > 75 %

Error Rate

Precision = $Q_{correct}$ / $Q_{answered}$



PRECISE made no mistakes on semantically tractable questions

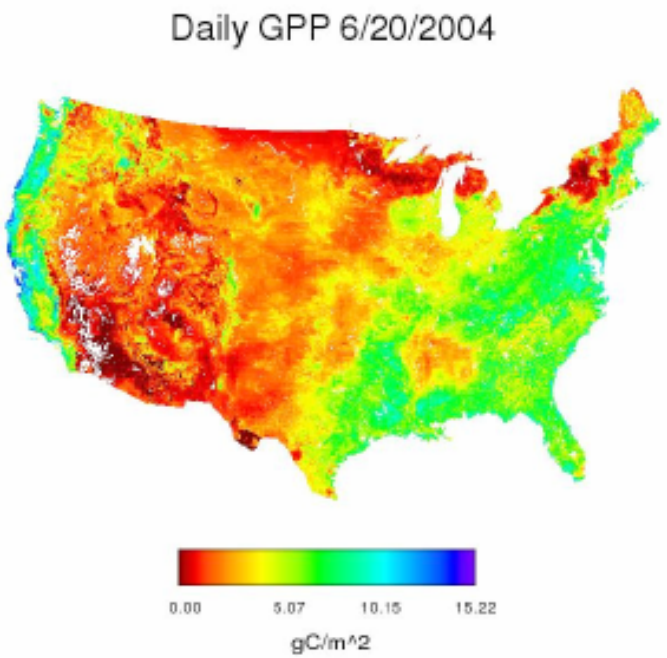
Ames Research Center

http://ecocast.arc.nasa.gov

ECOLOGICAL FORECASTING

Monitoring, Modeling, and Forecasting the Impacts of Climate Variability and Change on Ecosystems

Daily Ecocast



< previous | current > more images & data

What is Ecocasting?

Ecological forecasting (or 'ecocasting') is the prediction of ecosystem parameters. NASA Ames is developing advanced computing technologies for converting massive streams of satellite remote sensing data into ecocasts that are easy to read and use.

NASA Ames, UWF IHMC, CMU, CSUMB, UMT, UW, and Fetch Technologies are collaborating to develop a distributed computing architecture for the production of ecocasts from satellite remote sensing data and other ancillary data sources. Applications of the Ecocast technology include fire forecasting, crop quality forecasting, snowpack and flood monitoring, and identification of anomalies in the carbon cycle and other biospheric processes.

News

Daily updates of biospheric parameters are now available. See below for a selection of available parameters. Or download data and images here.

Nowcasts & Forecasts

- Meteorology
- Hydrology
- Carbon Cycle

Related Work (this session)

- GENESIS
  - SciFlo similar to JDAF
  - Doppler planner could be used to generate SciFlow “plans”
- IDACT
  - Automatic data type transformations
  - Restricted form of planning for dataflow generation
- DISCOVER
  - Transparent data access
  - Use of ESML
- GeoBrain
  - A geo-tree is a dataflow plan
  - No support (yet) for automatic generation

Other Related Work

- MVP, COLLAGE
  - Scientific image processing, human in the loop
  - HTN representation, less need for precise causal representation
  - No causal reasoning about data, metadata generation
- Chimera
  - Data tracking, but no support for causal reasoning
- Amphion, AutoBayes
  - Program synthesis using theorem proving
  - More expressiveness than needed for many DP problems
- Internet Softbot
  - Information gathering and changes to world.
  - Could never handle Unix pipes

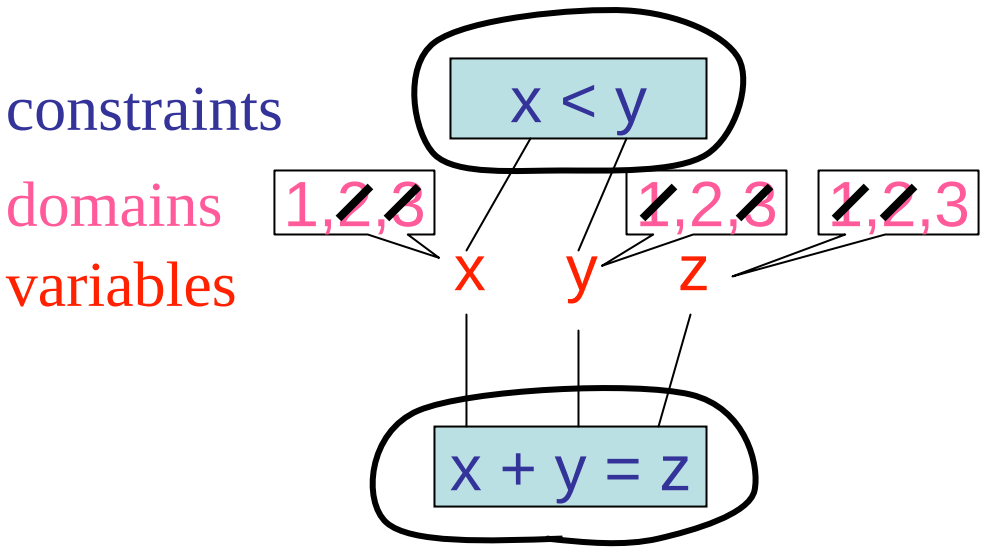
Status

- Website: http://ecocast.arc.nasa.gov
- Distributed real-time generation of NPP, GPP, soil moisture, and evapotranspiration for continental U.S
- Planner fully integrated with JDAF
- NLI integration in progress
- IDE (Expert user interface)
  - Domain editor with syntax/error highlighting
  - Graphical viewer/editor for plans, constraint network
- Web interface for planner
  - Bare-bones implementation complete
  - Work in progress

Ongoing & Future Work

- Daily phenological monitoring
- Invasive species tracking
- Data visualization
- PRECISE support for TOPS queries
- Translation of PRECISE queries into planner goals

Constraint Satisfaction



propagation